

# Deep Learning for Genome Annotation



Project Title	Deep Learning for Genome Annotation
Project Summary	Implement the latest Machine Learning approaches on HPC using millions of DNA sequences generated via Next Generation Sequencing platforms to decode the genome-level features of non-model marine invertebrates.
Country	United States

## Project Description

Multiple machine Learning approaches will be used to analyze genomics, transcriptomes and proteomics data in non-model invertebrates. The goal will be to identify the different regulatory elements in a eukaryotic genome to allow genomic annotation of various features in the noncoding regions, which make up the bulk of all genomes. Due to the size of the data processed and extent of the computational analysis to be performed, most of the work will be complete on our HPCC cluster HYDRA.

## Required Skills or Interests

Skill(s)

Coding

Data analysis

Data visualization

## Additional Information

Demonstrated data analysis skills:  
Proficient programing in Python and R  
Linux  
High Performance computing  
Machine learning algorithm and implementation  
Knowledge of Deep Learning and Support vector machines

Familiarity with auto encoders, tensor flow and caffe (preferred)

Parallel programming

A basic background in biology is preferred, especially related to the central dogma of molecular biology "DNA makes RNA and RNA makes protein" as well as the concept of epigenomics

Additional information:

Potential candidates must be able to work independently and conduct regular troubleshooting

Weekly or biweekly meetings (and others as needed) will be conducted to maintain team focus

Excellent communication skills are a must

## **Language Requirements**

*None*